



```

AC P09799;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT globulin (vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -
DR PFAM: PF00546; Seedstore_7s; 1.
DR Seed storage protein: Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
FT SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;
SQ
Query Match 76.7%; Score 188; DB 1; Length 605;
Best Local Similarity 75.0%; Pred. No. 2.5e-13;
Matches 30; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
OY 3 ROFOECQCHQHOEORPEKKQCCVRECKREKYQENPNGER 42
Db 120 KQKRECGQRCQMGQEQRPERRKQCCVRECKREKYQEDPWKGER 159
RESULT 3
TRX_DROVI STANDARD; PRT; 3828 AA.
ID AC 024742;
DT 01-OCT-2000 (rel. 40, Created)
DT 01-OCT-2000 (rel. 40, Last sequence update)
DE 01-OCT-2000 (rel. 40, Last annotation update)
DE TRITHORAX PROTEIN.
GN TRX.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE:96100387; PubMed:8555104;
RA Tillib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
RT between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF TRITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.

```

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -1- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 PHD-FINGER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z50038; CAA90349.1; -
DR HSSP: P04002; 1WPA.
DR FLYBASE: FBgn0014844; Dvir|trx.
DR INTERPRO: IPR001214; -
DR PFAM: PF00628; PHD; 2.
DR PFAM: PF00856; SET; 1.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator.
FT DOMAIN 1251 1334 PHD 1.
FT DOMAIN 1335 1380 PHD 2.
FT DOMAIN 1408 1469 PHD 3.
FT DOMAIN 1708 1767 PHD 4 (ATYPICAL).
FT DOMAIN 1768 1818 PHD 5 (ATYPICAL).
FT DOMAIN 3701 3810 SET DOMAIN.
FT DOMAIN 28 41 POLY-ALA.
FT DOMAIN 66 71 POLY-ASP.
FT DOMAIN 160 164 POLY-ASP.
FT DOMAIN 173 182 POLY-ALA.
FT DOMAIN 221 228 POLY-GLN.
FT DOMAIN 243 251 POLY-ALA.
FT DOMAIN 253 258 POLY-THR.
FT DOMAIN 292 296 POLY-ALA.
FT DOMAIN 296 306 POLY-ASP.
FT DOMAIN 1072 1075 POLY-GLU.
FT DOMAIN 2483 3271 GLN-RICH.
FT DOMAIN 3333 3339 POLY-ASP.
FT SEQUENCE 3828 AA; 413721 MW; 32059CF303A3C504 CRC64;
SQ
Query Match 30.2%; Score 74; DB 1; Length 3828;
Best Local Similarity 38.2%; Pred. No. 2.1;
Matches 13; Conservative 14; Mismatches 7; Indels 0; Gaps 0;
OY 2 QROFOECQCHQHOEORPEKKQCCVRECKREKYOE 35
Db 2995 QOEPEOEQCHLHQOQOQOQOQOQOQOQOQOQOQO 3028
RESULT 4
ALL1_SINL STANDARD; PRT; 127 AA.
ID AC P15322;
DT 01-APR-1990 (rel. 14, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DE 15-DEC-1998 (rel. 37, Last annotation update)
DE ALLERGEN SIN A I, SMALL AND LARGE CHAINS (SIN A I).
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Sinapis.
RN [1]
RP SEQUENCE.
RC TISSUE=SEED;
RA MEDLINE:89030681; PubMed:3181153;
RA Menendez-Arias L., Moneo I., Dominguez J., Rodriguez R.;
RT "Primary structure of the major allergen of yellow mustard (Sinapis
RT alba L.) seed, Sin a I.";
RL Eur. J. Biochem. 177:159-166(1988).

```

```
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- DISEASE: ALLERGIC DISEASE, CABBAGE ALLERGY.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR PIR: S01792; S01792.
DR PIR: S01791; S01791.
DR INTERPRO: IPR000617; -.
DR PFAM: PF01631; Seedstore_2s; 1.
DR PRINTS: PR00496; NAPIIN.
KM Allergen; Seed storage protein.
FT CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 127 LARGE CHAIN.
FT VARIANT 6 6 R -> G.
FT SEQUENCE 127 AA; 14180 MW; 4CD920284F04EEFD CRC64;

Query Match 29.0%; Score 71; DB 1; Length 127;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 7; Mismatches 8; Indels 22; Gaps 3;

OY 2 QROFOE-----COQHCHQO-----EQREPKKQCCVRCRCREKYQENP 37
Db 11 RKEFQQAQHLRACQOWLHKQAMQSGSPQGPQGRPLLQCC---CNELHQEPP 62

RESULT 5
ALL_BRAJU STANDARD; PRT; 129 AA.
AC P80207; P80215;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ALLERGEN BRA J 1-E, SMALL AND LARGE CHAINS (BRA J 1).
OS Brassica juncea (leaf mustard) (Indian mustard).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Brassica.
RN [1]
RP TISSUE=SEED.
RC MEDLINE=93356721; PubMed=7688955;
RX Monsalve R.I., Gonzalez de la Pena M.A., Menendez-Arias L.,
RA Lopez-Otin C., Villalba M., Rodriguez R.;
RT "Characterization of a new oriental-mustard (Brassica juncea)
allergen, Bra j 1E: detection of an allergenic epitope.";
RL Biochem. J. 293:625-632(1993).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY TWO DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR INTERPRO: IPR000617; -.
DR PFAM: PF01631; Seedstore_2s; 1.
DR PRINTS: PR00496; NAPIIN.
KM Allergen; Seed storage protein.
FT CHAIN 1 37 SMALL CHAIN.
FT NON_CONS 37 38
FT CHAIN 38 129 LARGE CHAIN.
FT VARIANT 6 6 R -> I.
FT VARIANT 20 20 F -> K.
FT SEQUENCE 129 AA; 14644 MW; D6F28E03F62B08F8 CRC64;

Query Match 29.0%; Score 71; DB 1; Length 129;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 7; Mismatches 8; Indels 22; Gaps 3;

OY 2 QROFOE-----COQHCHQO-----EQREPKKQCCVRCRCREKYQENP 37
Db 10 RKEFQQAQHLRACQOWLHKQAMQSGSPQGPQGRPLLQCC---CNELHQEPP 61
```

```
RESULT 6
ITRY_SINAR STANDARD; PRT; 130 AA.
AC P38057;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE TRYPSIN INHIBITOR (TISA).
OS Sinapis arvensis (Charlock).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Sinapis.
RN [1]
RP TISSUE=SEED.
RC MEDLINE=94350545; PubMed=8070965;
RX Svendsen I.B., Nicolova D., Goshv I., Genov N.;
RT "Primary structure, spectroscopic and inhibitory properties of a two-
chain trypsin inhibitor from the seeds of charlock (Sinapis arvensis
L), a member of the napin protein family.";
RT Int. J. Pept. Protein Res. 43:425-430(1994).
RL -1- FUNCTION: INHIBITS TRYPSIN WITH A KI OF 7 X 10(-6) M.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR INTERPRO: IPR000617; -.
DR PFAM: PF01631; Seedstore_2s; 1.
DR PRINTS: PR00496; NAPIIN.
KM Seed storage protein; Multigene family; Polymorphism.
FT CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 130 LARGE CHAIN.
FT VARIANT 32 32 R -> M.
FT VARIANT 73 73 A -> S.
FT VARIANT 77 77 K -> R.
FT VARIANT 81 81 Q -> R.
FT VARIANT 87 87 H -> Q.
FT VARIANT 89 89 Q -> H.
FT VARIANT 91 91 G -> Q.
FT VARIANT 97 97 E -> M.
FT VARIANT 98 98 I -> V.
FT VARIANT 99 99 R -> S.
FT VARIANT 106 106 T -> K.
FT VARIANT 123 123 N -> Q.
FT VARIANT 124 124 K -> G.
FT VARIANT 126 126 M -> V.
FT VARIANT 53 57 MISSING (IN FORM II).
FT SEQUENCE 130 AA; 14682 MW; EC02E4B26D18D0F2 CRC64;

Query Match 29.0%; Score 71; DB 1; Length 130;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 7; Mismatches 8; Indels 22; Gaps 3;

OY 2 QROFOE-----COQHCHQO-----EQREPKKQCCVRCRCREKYQENP 37
Db 11 RKEFQQAQHLRACQOWLHKQAMQSGSPQGPQGRPLLQCC---CNELHQEPP 62

RESULT 7
LP61_EIMTE STANDARD; PRT; 255 AA.
ID LP61_EIMTE
AC P15714;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE ANTIGEN LPMC-61 (FRAGMENT).
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
RN [1]
RP TISSUE=SPOROZOITE.
RC
```

```

RA MEDLINE=90348718; PubMed=2200963;
RX Ko C., Smith C.K. II, McDonnell M.;
RT "Identification and characterization of a target antigen of a
RT monoclonal antibody directed against Eimeria tenella merozoites.";
RL Mol. Biochem. Parasitol. 41:53-64(1990).
CC -!- FUNCTION: UNKNOWN. THE GLN-RICH TANDEM REPEATS MAY BE IMPORTANT
CC FOR AN UNKNOWN ASPECT OF THE PARASITIC LIFE CYCLE. MAY BE AN
CC IMPORTANT IMMUNOGEN.
CC -!- SUBUNIT: MAY BE COVALENTLY LINKED BY DISULFIDE BONDS TO OTHER
CC POLYPEPTIDES TO FORM THE 80 KDA ANTIGEN.
CC -!- DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES THROUGHOUT THE
CC SPOROULATION OF THE OOCYSTS AND IN THE SPOROZOITES FOLLOWING
CC EXCYSTATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL, M30933; AAA29079.1; -.
DR PIR, A60637; A60637.
KW Antigen; Sporozoite; Repeat; Sporulation.
FT FT DOMAIN 1 1
FT FT 18 210 12 X APPROXIMATE TANDEM REPEATS, GLN-
FT FT RICH.
FT REPEAT 18 48 1.
FT REPEAT 49 57 2.
FT REPEAT 58 65 3.
FT REPEAT 66 78 4.
FT REPEAT 79 90 5.
FT REPEAT 91 103 6.
FT REPEAT 104 140 7.
FT REPEAT 141 152 8.
FT REPEAT 153 164 9.
FT REPEAT 165 172 10.
FT REPEAT 173 192 11.
FT REPEAT 193 210 12.
FT FT NON_TER 255
SQ SEQUENCE 255 AA; 31267 MW; 8C5E6055FFFC2DB3 CRC64;
OY 1 SORQRFQEQQHCHQO-----EQREKKQCCRECKREKQENPMW 38
DB 122 SQQQQLQCCGQQQQQQQLQQQWSEQQQQQQQQQQWPEQPEQQQQQQQQW 165

RESULT 8
FTX1_HUMAN
AC P54253;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ATAXIN-1 (SPINOCEREBELLAR ATAXIA TYPE 1 PROTEIN).
GN SCA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREBELLUM, AND BRAIN;
RX MEDLINE=95038838; PubMed=7951322;
RA Banfi S., Servadio A., Chung M.-Y., Kwiatkowski T.J. Jr., McCall A.E.,
RA Duvick L.A., Shen Y., Roth E.J., Orr H.T., Zoghbi H.Y.;
RT "Identification and characterization of the gene causing type 1
RT spinocerebellar ataxia.";
```

RL Nat. Genet. 7: 513-519(1994).  
CC -I- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY  
CC ALTERNATIVE SPLICING.  
CC -I- TISSUE SPECIFICITY: WIDELY EXPRESSED THROUGHOUT THE BODY.  
CC -I- POLYMORPHISM: THE POLY-GLN REGION OF SCAl IS HIGHLY POLYMORPHIC  
CC (9 TO 39 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO  
CC ABOUT 40-81 REPEATS IN SCAl PATIENTS. LONGER EXPANSIONS RESULT IN  
CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE  
CC DISEASE.  
CC -I- DISEASE: DEFECTS IN SCAl ARE THE CAUSE OF SPINO CEREBELLAR ATAXIA  
CC TYPE I; ALSO KNOWN AS OLIVOPONTOCEREBELLAR ATROPHY I (OPCA I).  
CC SCAl IS AN AUTOSOMAL DOMINANT NEURODEGENERATIVE DISORDER  
CC CHARACTERIZED BY PROGRESSIVE NEURONAL LOSS IN THE CEREBELLUM,  
CC BRAIN STEM AND SPINO CEREBELLAR TRACTS. CLINICAL FEATURES ARE  
CC CEREBELLAR ATAXIA, DYSPARTHRIA, OPHTHALMOPARESIS, MUSCLE WASTING AND  
CC NEUROPATHY. ONSET OF THE DISEASE USUALLY OCCURS IN THE THIRD OR  
CC FOURTH DECADE OF LIFE AND DEATH OCCURS TEN TO TWENTY YEARS LATER.

--

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL; X79204; CAA55793.1; -.  
DR MIM; 164400; -.  
KW Polymorphism; Triplet repeat expansion; Alternative splicing.  
FT DOMAIN 197 226 POLY-GLN.  
SQ SEQUENCE 816 AA; 87051 MW; DA9BA5DBA23D0777 CRC64;

```

Query Match          26.9%; Score 66; DB 1; Length 816;
Best Local Similarity 52.2%; Pred. No. 3.8;
Matches 12; Conservative 7; Mismatches 4; Indels 0; Gaps 0.

OY      2  OROFECOOHCHQOEORPEKKOO 24
          | : | : | | | | : : | |
Db      200  OQQQQQQQQQHQQHQQQQQQQQQQO 222

RESULT          9
YAK1_YEAST
AC      PI4680; STANDARD; PROT; 807 AA.
DT      01-APR-1990 (Rel. 14, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      PROTEIN KINASE YAK1 (EC 2.7.1.-).
GN      YAK1 OR YUL141C OR J0652.
GC      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC      Saccharomycetaceae; Saccharomyces.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90108683; Pubmed=2558053;
RA      Garrett S., Broach J.;
RT      "Loss of Ras activity in Saccharomyces cerevisiae is suppressed by
RT      disruptions of a new kinase gene, YAK1, whose product may act
RT      downstream of the GMP-dependent protein kinase."
RL      Genes Dev. 3:1336-1348(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      STRAIN=S288C / FY1679;
RX      MEDLINE=964088771; Pubmed=8813765;
RA      Katsoulou C., Tzermita M., Tavernarakis N., Alexandraki D.;
RT      "Sequence analysis of a 40.7 kb segment from the left arm of yeast
RT      chromosome X reveals 14 known genes and 13 new open reading frames
RT      including homologues of genes clustered on the right arm of
RT      chromosome XI."
RL      Yeast 12:787-797(1996).
CC      -1- FUNCTION=THIS PROTEIN OCCURS IN REVERTANT RAS/CAMP MUTANTS

```

```
CC YEAST AND MAY BE INVOLVED IN CELL-CYCLE REGULATION.
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAM/DYRK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16056; CAA34192.1; -
DR EMBL: X87371; CAA60814.1; -
DR EMBL: Z49417; CAA89437.1; -
DR PIR: A32582; A32582.
DR HSSP: P24941; IAO1.
DR SGD: S0003677; YAK1.
DR INTERPRO: IPR000719; -.
DR INTERPRO: IPR002290; -.
DR PFAM: PF00069; PKinase_2.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST_1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM_1.
KW Transferrase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation.
KT DOMAININ 56 GLN-RICH.
FT DOMAIN 369 704 PROTEIN KINASE.
FT NP_BIND 375 383 ATP (BY SIMILARITY).
FT BINDING 398 398 ATP (BY SIMILARITY).
FT ACT_SITE 496 496 BY SIMILARITY.
FT MOD_RES 127 127 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 206 206 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 240 240 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 295 295 PHOSPHORYLATION (PROBABLE).
SQ SEQUENCE 807 AA; 91245 MW; EDB7C56FFA35E056 CRC64;

Query Match 26.3%; Score 64.5; DB 1; Length 807;
Best Local Similarity 34.0%; Pred. No. 5.4;
Matches 16; Conservative 8; Mismatches 10; Indels 13; Gaps 2;

QY 1 SQQRQECQHCHQDEORPEKKQ-----CYRECKEKTKQENPMGER 42
Db 58 SQSRPOQDHONHHOQOOOQOOOQONSQCF-----VNPWNEEK 96

RESULT 10
2SSE_BRANA STANDARD; PRT; 186 AA.
AC AC P09893;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE MAPIN EMBRYO SPECIFIC PRECURSOR (1.7'S SEED STORAGE PROTEIN).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Brassica.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308225; PubMed=3040733;
RA Scofield S.R., Crouch M.L.;
RT "Nucleotide sequence of a member of the napin storage protein family
RT from Brassica napus.";
RL J. Biol. Chem. 262:12202-12208(1987).
CC -I- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC ITS MATURATION.
CC -I- SUBUNIT: THE MAJURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -I- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
CC -I- DEVELOPMENTAL STAGE: EMBRYO.
```

```

CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J02782; AAA33007.1; -.
CC PIR: A29802; A29802.
CC INTERPRO: IPR000617; -.
CC PFMAM; PF01631; Seedstore_2s; 1.
CC PRINTS; PR00496; NADIN.
CC Seed storage protein; Signal; Multigene family.
CC SIGNAL 1 21
CC PROPEP 22 38
CC CHAIN 39 76 SMALL CHAIN.
CC PROPEP 77 97
CC CHAIN 98 186 LARGE CHAIN.
CC SEQUENCE 186 AA; 21013 MW; 9CAFE63D84B160AB3 CRC64;
CC
CC Query Match 26.1%; Score 64; DB 1; Length 186;
CC Best Local Similarity 24.7%; Pred. NO. 1.6;
CC Matches 18; Conservative 7; Mismatches 8; Indels 40; Gaps 3;
CC
CC QY 2 OROPE-----COOHCHQO-----EGREKKQO 24
CC Db 48 RKEPQAOHLRACQOOWLHKQAMQPGSGSPKWTLDGEFDVEDVENQOQGPORRPPPOQ 107
CC QY 25 CVRECEKRYQENP 37
CC 1 1 1 1 1 1
CC Db 108 C--CNELHQDEP 117
CC
CC RESULT 11
CC HOG1 HORVU
CC ID HOG1 HORVU STANDARD; PRT; 305 AA.
CC AC P12990;
CC DT 01-NOV-1990 (Rel. 16, Created)
CC DT 01-NOV-1990 (Rel. 16, Last sequence update)
CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
CC DE GAMMA-HORDEIN 1 PRECURSOR.
CC OS Hordeum vulgare (Barley).
CC OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
CC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP STRAIN-CV, CARINA;
CC RC Cameron-Mills V., Brandt A.;
CC RA "A gamma-hordein gene."
CC RT Plant Mol. Biol. 11:449-461(1988).
CC CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (AS GLOBULES) AND VACUOLAR
CC (AS PROTEIN BODIES).
CC CC -1- TISSUE SPECIFICITY: DEVELOPING ENDOSPERM.
CC CC -1- DOMAIN: SULFUR RICH HORDEIN WHICH POSSES AN N-TERMINAL HALF
CC COMPOSED OF PROLINE-GLUTAMINE BLOCKS ORGANIZED IN REPEATING UNITS
CC AND A C-TERMINAL HALF WHERE THE REPEATS ARE DISPERSED AND LESS
CC CONSERVED.
CC CC -1- SIMILARITY: TO GAMMA-GLIADIN FROM WHEAT, AND A GAMMA-SECALIN FROM
CC RYE, AND LESS TO A BI HORDEIN FROM BARLEY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X13508; CAA31861.1; -.

```

DR EMBL: M36378; AAA32955.1; -  
 DR PIR: S08312; S08312.  
 DR INTERPRO: IPR000528; -  
 DR INTERPRO: IPR001954; -  
 DR PFAM: PF00279; LRP; 1.  
 DR PRINTS: PR00208; GLINDGUTEN.  
 DR Seed storage protein; Signal; Multigene family.  
 KW SIGNAL  
 FT SIGNAL 1  
 FT CHAIN 20 305 GAMMA-HORDEIN 1.  
 FT SEQUENCE 305 AA; 34737 MW; 6D803B53EFH24AD CRC64;

Query Match 26.1%; Score 64; DB 1; Length 305;  
 Best Local Similarity 35.6%; Pred. No. 2.5;  
 Matches 16; Conservative 12; Mismatches 11; Indels 6; Gaps 2;

OY 1 SORFQECQCHQOEOR--PEKKQCVRECKEYK---QENWR 39  
 DB 48 S00QPPQ0PPQPPQPPQ0Q0CLOQPHQPPQTPQEPFR 92

## RESULT 12

SNF5\_YEAST STANDARD; PRT; 905 AA.

ID P18480;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE TRANSCRIPTION REGULATORY PROTEIN SNF5 (SWI/SNF COMPLEX COMPONENT SNF5)

GN SNF5 OR YTE4 OR SWI10 OR YBR289W OR YBR2036.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

OC Saccharomycetaceae; Saccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-MCY;

RX MEDLINE-91042489; PubMed-2233708;

RA Laurent B.C., Tittel M.A., Carlson M.;

RT "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and

proline-rich transcriptional activator that affects expression of a

RT broad spectrum of genes.";

RL Mol. Cell. Biol. 10:5616-5625(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C;

RX MEDLINE-94378722; PubMed-8091861;

RA Holmstrom K., Brandt T., Kallio T.;

RT "The sequence of a 32,420 bp segment located on the right arm of

RT chromosome II from Saccharomyces cerevisiae.";

RL Yeast 10:547-562(1994).

CC -I- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF

CC OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE

CC BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.

CC -I- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR

CC COMPLEX.

CC -I- SUBCELLULAR LOCATION: NUCLEAR.

CC -I- SIMILARITY: BELONGS TO THE SNF5 FAMILY.

CC -----  
 CC This SNF5-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL: M36482; AAA35062.1; -

DR EMBL: X76053; CAA53652.1; -

DR EMBL: 236158; CAA85254.1; -

DR PIR: S44551; KGBRYS5.

DR PIR: S39145; S39145.

DR SGD: S0000493; SNF5.  
 KW Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 31 270 GLN-RICH.  
 FT DOMAIN 72 132 PRO-RICH.  
 FT DOMAIN 272 324 PRO-RICH.  
 FT DOMAIN 489 588 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 714 882 PRO-RICH.  
 FT DOMAIN 755 798 ARG/LYS-RICH (BASIC).  
 FT CONFLICT 564 564 E -> D (IN REF. 1).  
 FT SEQUENCE 905 AA; 102557 MW; A287B4A648DD1A35 CRC64;

Query Match 26.1%; Score 64; DB 1; Length 905;  
 Best Local Similarity 35.3%; Pred. No. 6.7;  
 Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

OY 2 QRFQECQCHQOEORPEKKQCVRECKEYKOE 35  
 DB 222 QKQ00000H00000000000000000000000000 255

## RESULT 13

TRHY\_HUMAN STANDARD; PRT; 1898 AA.

ID Q07283;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE TRICHOHYALIN.

GN THY OR TRHY OR THL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-93280194; PubMed-7685034;

RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,

RA Steinhert P.M.;

RT "The structure of human trichohyalin. Potential multiple roles as a

RT functional EF-hand-like calcium-binding protein, a cornified cell

RT envelope precursor, and an intermediate filament-associated (cross-

RT linking protein.";

RL J. Biol. Chem. 268:12164-12176(1993).

RN [2]

RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.

RX MEDLINE-93315897; PubMed-7686953;

RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinhert P.M.;

RT "Trichohyalin: a structural protein of hair, tongue, nail, and

RT epidermis.";

RL J. Invest. Dermatol. 101:655-715(1993).

CC -I- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES

CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE

CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR

CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY

CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER

CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN

CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN

CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL

CC DIFFERENTIATION.

CC -I- SUBUNIT: MONOMER (PROBABLE).

CC -I- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS

CC THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN

CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).

CC -I- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF

CC THE EPIDERMIS.

CC -I- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND

CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST

CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS

CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STANDARD

CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS

CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.

CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN

CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG

CC DIFFERENT SPECIES.  
CC -I- PM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE  
CC PROBABLY CONVERTED TO CITULLINES BY PEPTIDYLARGININE DEIMIDASE.  
CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100  
CC FAMILY.  
CC -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
-----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC CC  
CC DR EMBL; L09190; AAA6582.1; .  
CC DR PIR; A45973; A45973.  
CC DR HSSP; P02633; IBOC.  
CC DR MIM; 190370; .  
CC DR INTERPRO; IPRO01751; .  
CC DR INTERPRO; IPRO02048; .  
CC DR PFAM; PF01023; S\_100; 1.  
CC DR PFAM; PF00036; efhand; 1.  
CC DR PROSITE; PS00018; EF\_HAND; 1.  
CC DR PROSITE; PS00303; S100\_CABP; 1.  
KW Repeat; Calcium-binding.  
FT DOMAIN 1 91  
FT CA\_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).  
FT CA\_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).  
FT DOMAIN 314 390 6 X 13 AA TANDEN REPEATS OF  
R-R-E-Q-E-E-R-E-Q-Q-L.  
1-1 (APPROXIMATE).  
1-2 (APPROXIMATE).  
1-3 (APPROXIMATE).  
1-4.  
1-5.  
1-6.  
9 X 6 AA TANDEN REPEATS OF R-R-E-Q-Q-L.  
2-1.  
2-2.  
2-3.  
2-4.  
2-5.  
2-6.  
2-7.  
2-8.  
2-9.  
9 X 28 AA APPROXIMATE TANDEN REPEATS.  
8 X 30 AA TANDEN REPEATS.  
4-1.  
4-2.  
4-3.  
4-4.  
4-5.  
4-6.  
4-7.  
4-8.  
23 X 26 AA APPROXIMATE TANDEN REPEATS.  
F->L (IN REF. 2).  
QEHROYR->RSFGSTG (IN REF. 2).  
O->K (IN REF. 2).  
V->G (IN REF. 2).  
FT CONFLICT 1857 1880  
FT CONFLICT 1880 1880  
FO SEQUENCE 1898 AA: 247219 MW; AV4B5947EB62E31D CRC64;

```

Query Match      26.1%  Score 64; DB 1; Length 1898;
Best Local Similarity 31.8%  Pred. NO, 13;
Matches 14; Conservative 14; Mismatches 12; Indels 4; Gaps 1
OY 2 QROFOECQOHCHQOEDORPEKKOOCVARECKRYQ---ENPARG 41
      : : : : : : : : : : : : : : : : : : : : : :
Ob 1117 KRRROERERORCRREKEELIOOEPLLREERERKRROELERQREE 1160
      : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT_14
2SSI_BRANA
ID 2SSI_BRANA STANDARD: PRT: 110 AA.
AC P24565;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE NAPIIN IA AND IB SMALL CHAIN AND LARGE CHAINS.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Brassica.
RN [1]
RP SEQUENCE.
RC TISSUE=SEED;
RX MEDLINE=92111741; PubMed=1765156;
RA Monsalve R.I., Lopez-Olin C., Villalba M., Rodríguez R.;
RT "A new distinct group of 2 S albumins from rapeseed. Amino acid
sequence of two low molecular weight napins."
RL FEBS Lett. 295:207-210(1991).
CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPIINS ARE ONE OF THE
CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC ITS MATURATION.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- MISCELLANEOUS: NAPIIN IA AND IB ARE MINOR COMPONENT OF SEED 2S
CC ALBUMIN.
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF NAPIIN IA.
CC PIR: S20350; S20350.
DR INTERPRO: IPR000617;
DR PfAM: PF01631; seedstore_2s; 1.
DR PRINTS: PR00496; NAPIIN.
KW Seed storage protein.
FT CHAIN 1 SMALL CHAIN.
FT NON_CONS 31 32
FT CHAIN 32 110
FT VARIANT 31 31
FT VARIANT 37 37
FT VARIANT 107 110 MISSING (IN NAPIIN IB).
FT SEQUENCE 110 AA; 12691 MW; 342938ADA2C1E995 CRC64;

Query Match 25.9%; Score 63.5; DB 1; Length 110;
Best local Similarity 33.3%; Pred. No. 1.2;
Matches 17; Conservative 7; Mismatches 8; Indels 19; Gaps 3;
QY 2 QROFOE-----COQHCHQ-----EORPEKKQOCVRECREKYOEN 36
11:11: 111 11 : : : 1 1 111:
6 QREFOQEQHTRACQOWITRQQLAGSPFGSGPQEGWMLRQC---CNELYQED 53

RESULT_15
2SS2_BRANA
ID 2SS2_BRANA STANDARD: PRT: 178 AA.
AC P01090;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE NAPIIN 2 PRECURSOR (1.7S SEED STORAGE PROTEIN).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Brassica.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87308224; PubMed=3624251;
RA Josefsson L.-G., Lenman M., Ericson M.L., Raak L.;
RT "Structure of a gene encoding the 1.7 S storage protein, napin, from
J. Biol. Chem 262:12196-12201(1987).

```

```
RN [2]
RP REVISIONS.
RA Josefsson L.-G.;
RL Submitted (JUL-1987) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87033665; PubMed=3771543;
RA Ericson M.L., Roedin J., Lenman M., Glimelius K., Josefsson L.-G.,
RA Rask L.;
RT "Structure of the rapeseed 1.7 S storage protein, napin, and its
RT precursor";
RL J. Biol. Chem. 261:14576-14581(1986).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, TOWER;
RX MEDLINE=84113267; PubMed=6689334;
RA Crouch M.L., Tenberge K.M., Simon A.E., Perl R.;
RT "cDNA clones for Brassica napus seed storage proteins: evidence from
RT nucleotide sequence analysis that both subunits of napin are cleaved
RT from a precursor polypeptide.";
RL J. Mol. Appl. Genet. 2:273-283(1983).
CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC ITS MATURATION.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K01545; AAA33006.1; -
DR EMBL: J02586; AAA32997.1; -
DR EMBL: J02798; AAA87348.1; -
DR PIR: A01329; NMRP2.
DR PIR: A29801; A29801.
DR PIR: A25997; A25997.
DR INTERPRO: IPR000617; -
DR PFAM: PF01631; SeedStore_25; 1.
DR PRINTS: PR00496; NAPIN.
KW Seed storage protein; Signal; Multigene family.
FT SIGNAL 1 21
FT PROPEP 22 38
FT CHAIN 39 74 SMALL CHAIN.
FT PROPEP 75 94
FT CHAIN 95 175 LARGE CHAIN.
FT CONFLICT 37 37 D -> N (IN REF. 4).
FT CONFLICT 76 76 S -> N (IN REF. 4).
SQ SEQUENCE 178 AA: 20104 MW: 734E561971B539FF CRC64;
```

Query Match 25.9%; Score 63.5; DB 1; Length 178;  
Best Local Similarity 25.7%; Pred. No. 1.8;  
Matches 18; Conservative 7; Mismatches 8; Indels 37; Gaps 3;

QY 2 QROFGE-----CQHCHQO-----EORPEKKQCVR 27  
:::11: 111 1:-1 :111 111  
DB 48 RKEFDQAQHLIRACQOQWLRQAMQSGGSPWTLDFEFDFEDMENPQGPQORPPLQQC-- 105

QY 28 ECREKYOENP 37  
| | :111 |  
DB 106 -CNEIHQEER 114

Search completed: March 1, 2001, 16:16:58  
Job time: 438 sec